GRAVITE Support for NOAA-20 VIIRS SDR Reprocessing

Wayne McCullough
wayne.mccullough@noaa.gov
DPES Sustainment Team

Overview

- Background
  - GRAVITE
  - PGEs
  - ADL
- Request
  - NOAA 20 VIIRS had unexpected conditions after launch
  - SDR team wanted to reprocess with new LUTs
- Process
  - Strong user interaction
- Results
  - First run
  - Second run
- Conclusion

GRAVITE IPS

- GRAVITE has *a lot* of data
  - 112 Million Files, 91 Million unique granules (SNPP and NOAA-20)
  - All RDRs since launch, 34 day rolling storage of other XDRs
  - 560 TB of data
- GRAVITE has available resources
  - Computer
    - Workstations (at GSFC L40)
    - ICF Servers: dedicated to remote access and compute tasks
    - PGE Servers: dedicated to automated processing
    - *All servers have direct access to data*
    - 1.1 PB of dedicated disk space for operational system
  - Tools
    - IDL, Matlab, Python, Redmine, PGEs, etc.
  - Support
    - Operators, Developers, Engineers, etc.
- It is there for the JPSS Cal-Val & Data Quality community to use

GRAVITE PGEs

- What is a PGE?
  - Product Generation Executable
  - Any code we automatically run against data for time periods
    - E.g., Run this analysis every hour when the data is available, etc.
- Initially used heavily for ground comparison
- Broader use now:
  - Instrument DQ Checks
  - Static plot generation
  - Data Preview Tile sets
  - Granulated Ancillary generation (uses ADL)
  - Reprocessing

GRAVITE PGEs

- Conditions for running a PGE:
  - Have rules defining the time periods (Execution Blocks)
    - Hourly, daily, etc.
    - Orbit
    - Custom lookup
  - Define input products
    - Input products may be optional or required.
    - A minimum number or maximum time gap may be set.
    - A geo-spatial area may be set.
    - E.g. *“If I have full coverage for VIIRS M7, the cloud mask, and the GEOs, run xxx for this hour”*
  - Automatically run PGE when conditions are met
    - Queue task
    - Execute code on available node
  - Selected Outputs are archived by GRAVITE

ADL Architecture

- Simplified reflection of IDPS architecture:
  - Processing Subsystem (PRO)
  - Data Management Subsystem (DMS)
  - No Ingest Subsystem (ING), No Data Delivery Subsystem (DDS), No Infrastructure Subsystem (INF)
    - Some functions replaced by ADL Toolkit

Request

- STAR VIIRS SDR team needed to regenerate SDRs from RDRs
  - Unexpected conditions shortly after launch
  - New Lookup Tables needed
  - Wanted to reprocess all NOAA-20 VIIRS Science RDRs (from launch November 2017 to end of February 2018)
- February 2018 DPES and STAR VIIRS SDR team met
  - Various approaches considered
    - All centered on many runs of ADL
  - SDR team needed more time to finalize LUTs
  - DPES team needed more time to test and refine ADL calls
- Goal:

Start Processing by August 2018

Process

- DPES dedicated three computers in GRAVITE for run
  - Each machine 24 core, 256 GB RAM
    - Dell PowerEdge R430 servers, with two Intel Xeon E5-2680v3 CPUs at 2.5GHz, eight 32GB RDIMM with Advanced ECC, and two Intel Ethernet X540 DP 10GBASE-T
  - Each machine to run a max of 16 ADL processes
- Set up ADL to run in a PGE
  - STAR VIIRS SDR team provided significant support:
    - patch to ADL to turn on compression
    - Testing and reference data
    - Final LUT package for reprocessing
- Runs as part of operational PGE system
  - No impact to current PGEs, only minor configuration changes needed
  - Reprocessing PGE delivered in GRAVITE v4.4

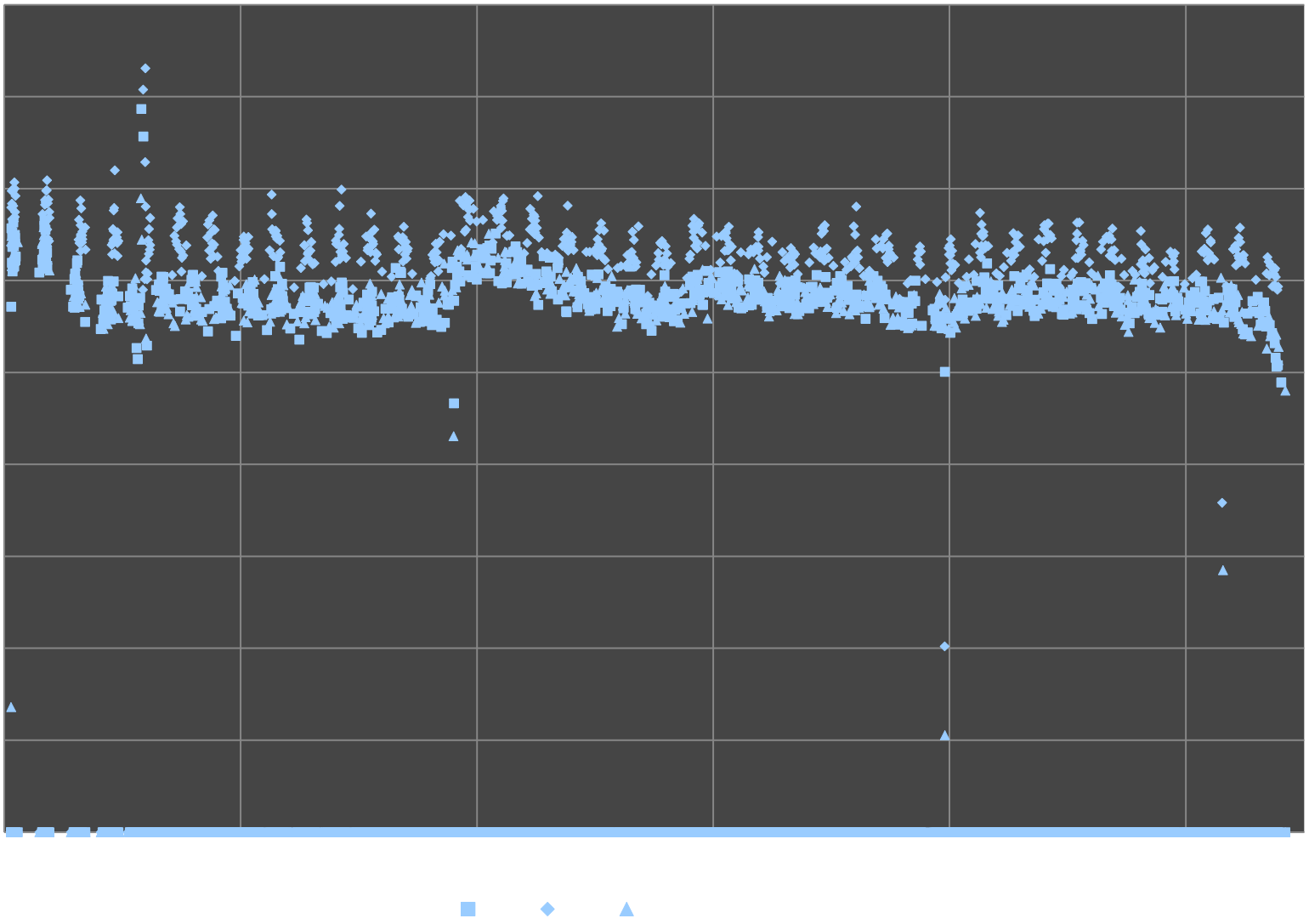
Results

- First Run start 2018-07-05
- Finish 2018-07-17
- VIIRS SDR team noted that about 1% of outputs were missing
  - ADL was not called properly to handle A2 Granules
  - A fix was developed
- Decided to re-run everything
  - Ensure all data was correctly processed.
  - Avoid duplicates.

Results

- Second Run start 2018-08-16 19:22
- Finish 2018-08-27 19:01
- Performance: ~85 days of data processed in 11 days
  - Average time to run each hour of data: 6 hours
  - Run 48 simultaneously across three computers
  - Net: ~*7.8x faster than real time*
  - If needed, we could parallelize it more
- Outputs: ~42 TB of data
  - Segmented into the **gvo** domain. (i.e., Not ops, pop, or int)
  - DPES will keep for 1 year
  - Available for all GRAVITE users

Time to Run ADL



Conclusion

- GRAVITE can use ADL to reprocess large amounts of data
  - ADL is a complex utility
  - Requires a bit of trial an error
- GRAVITE IPS PGE system can support large reprocessing runs
  - This is the first run of this nature we have tried
  - Overall, the GRAVITE IPS system did what it is supposed to do
- DPES can support JPSS Reprocessing via ADL
  - ADL available on ICF machines
  - Talk with DPES for larger runs

Contacts

- To subscribe to DQA alerts, contact:
  - ops-gravite-dpes-jpss@lists.nasa.gov

  (Subscribers need to have a GRAVITE account)
- New GRAVITE account request, contact:
  - Erica Handleman: erica.handleman@nasa.gov
- System access issues, contact:
  - gravite.service@noaa.gov
- DQA functions, contact:
  - dqst-dpes-jpss@lists.nasa.gov
- All other issues, contact:
  - ops-gravite-dpes-jpss@lists.nasa.gov